UTEC-CSc-81-020
Final Technical Report

Copy # 11

LEVEL III

AD A100629

NOISE SUPPRESSION METHODS FOR ROBUST SPEECH PROCESSING

Contractor: University of Utah
Effective Date: 2 January 1979
Expiration Date: 31 March 1981

Principal Investigator: Dr. Steven F. Boll
Telephone: (801) 581-8224

Sponsored by
Defense Advanced Research Projects Agency (DoD)
ARPA Order No. 3301
Monitored by Naval Research Laboratory
Under Contract No. N00173-79-C-0045

April 1981

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UTEC-CSC-81-020
2. GOVT ACCESSION NO.: AD-A100 629
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): NOISE SUPPRESSION METHODS FOR ROBUST SPEECH PROCESSING.
5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report, 1 Oct 78-31 Mar 81
6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Steven F. Boll, James Kajiya, James Youngberg, Tracy A. Petersen, H. Ravindra, William Done, B.V. Cox, Elaine Cohen
8. CONTRACT OR GRANT NUMBER(s): N00173-79-C-0045, ARPA Order-3341

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
University of Utah
Computer Science Department
Salt Lake City, Utah 84112

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Project: 76-RPA-3301

11. CONTROLLING OFFICE NAME AND ADDRESS:
Defense Advanced Research Project Agency (DoD)
1400 Wilson Boulevard
Washington, D.C. 22209

12. REPORT DATE: April 81
13. NUMBER OF PAGES: 48

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
Naval Research Laboratory
455 Overlook Ave. S.W.
Mail Code 2415-A.M.
Washington, D.C.
UTEC-CSC-81-020

15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Acoustic noise suppression in speech, Adaptive noise cancellation, LMS algorithm, Lattice gradient algorithm, Wiener Filtering, Short-time Fourier analysis, Waveform coding, Articulation rate change, Constant-Q Analyzer, Critical Band Diagnostic Rhyme Test, LPC-10, Perceptual Modeling, Pole-zero modeling, Speech activity detection, Splines

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Robust speech processing in practical operating environments requires effective environmental and processor noise suppression. This report describes the technical findings and accomplishments during the reporting period for the research program funded to develop real-time, compressed speech analysis-synthesis algorithms whose performance is invariant under signal contamination. Fulfillment of this requirement is necessary to insure reliable secure compressed speech transmission within realistic military command and control environments. Overall contributions resulting from this research

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404949 LB

program include the understanding of how environmental noise degrades narrow band, coded speech, development of appropriate real-time noise suppression algorithms, and development of speech parameter identification methods that consider signal contamination as a fundamental element in the estimation process. This report describes the research and results in the areas of noise suppression using the dual input adaptive noise cancellation articulation rate change techniques, spectral subtraction and a description of an experiment which demonstrated that the spectral subtraction noise suppression algorithm can improve the intelligibility of 2400 bps, LPC-10 coded, helicopter speech by 10.6 points. In addition summaries are included of prior studies in Constant-Q signal analysis and synthesis, perceptual modelling, speech activity detection, and pole-zero modelling of noisy signals. Three recent studies in speech modelling using the critical band analysis-synthesis transform and using splines are then presented. Finally a list of major publications generated under this contract is given.

TABLE OF CONTENTS

TABLE OF CONTENTS

SECTION I

SUMMARY OF PROGRAM

PROGRAM OBJECTIVES

To develop practical, low cost, real-time methods for suppressing noise which has been acoustically added to speech.

To demonstrate that through the incorporation of the noise suppression methods speech can be effectively analysed for narrow band digital transmission in practical operating environments.

SUMMARY OF TASKS AND RESULTS

INTRODUCTION

In Section II the key research efforts of the program are summarized. In Section IV, three recent technical papers are presented. Section V lists the major publications generated under this contract.

SECTION II

SUPPRESSION OF ACOUSTIC NOISE IN SPEECH USING SPECTRAL SUBTRACTION

I. INTRODUCTION

Background noise acoustically added to speech can degrade the performance of digital voice processors used for applications such as speech compression, recognition, and authentication [1]. The effects of background noise can be reduced by using noise-cancelling microphones, internal modification of the voice processor algorithms to explicitly compensate for signal contamination, or preprocessor noise reduction. Noise-cancelling microphones, although essential for extremely high noise environments such as the helicopter cockpit, offer little or no noise reduction above 1 kHz [1]. Techniques available for voice processor modification to account for noise contamination are being developed [4]. Preprocessor noise reduction offers the advantage that noise stripping is done on the waveform itself with the output being either digital or analog speech. Thus, existing voice processors tuned to clean speech can continue to be used unmodified. Also, since the output is speech, the noise stripping becomes independent of any specific subsequent speech processor implementation (it could be connected to a CCD channel vocoder or a digital LPC vocoder).

The objectives of this research were to develop a noise suppression technique, implement a computationally efficient algorithm, and test its performance in actual noise environments. The approach used was to estimate the magnitude frequency spectrum of the underlying clean speech by subtracting the noise magnitude spectrum from the noisy speech spectrum. The average

noise magnitude was measured during nonspeech activity. The noise suppressor is implemented using about the same amount of computation as required in a FFT convolution. It is tested on speech recorded in a helicopter environment. Its performance is measured using the Diagnostic Rhyme Test (DRT) [6].

SIGNAL II. ESTIMATION USING SPECTRAL SUBTRACTION [3],[4]

Signal x(i) digitized from a single microphone consists of the sum of speech Sp(i) and ambient acoustic noise n(i). It is assumed that the noise is locally stationary to the extent that average value of its spectral magnitude during speech activity is equal to that measured just prior to speech activity. Using these assumptions the spectral subtraction algorithm attempts to suppress the additive acoustic noise component n(i) from x(i) by the following steps:

1. Segment the noisy data into windowed analysis blocks of length M samples, x(i),i=0,1...,M-1.

2. Compute the N point DFT X(k) of data x(i).

3. Estimate the speech spectrum S(k) by subtracting the average noise spectral magnitude, B(k) = ave|N(k) |, calculated during non-speech activity, from |X(k) |:

$$S(k) = [|X(k)| - B(k)] \exp(j\, ARG[X(k)])$$

The motivation behind this approach is to subtract from the noisy speech spectrum, an estimate of the noise spectrum which is readily available. The magnitude of N(k) is replaced by its average value, B(k), and the phase of N(k) is replaced by the phase of X(k).

INTELLIGIBILITY AND QUALITY TESTING RESULTS ON SPECTRAL SUBTRACTION AND LPC-10 [5]

Experiment Definition

The data base consisted of a three-speaker Diagnostic Rhyme Test (DRT) list recorded in the RH-53 helicopter. This data base was processed by the real-time spectral subtraction algorithm as implemented on the Utah FPS-120B array processor. Audio tapes consisting of the original digital source and the spectral subtraction output were then sent to the Naval Research Laboratory, (NRL). Each of these tapes were processed through NRL's LPC-10 2400bps real-time bandwidth compression system, generating two more tapes: original digital source with LPC, and spectral subtraction output with LPC. Finally these four tapes were sent to Dynastat for intelligibility scoring.

Results

The total DRT score for each tape is:

Original Digitized Source	= 85.2
Spectral Subtraction Output	= 79.8
Original Digitized Source With LPC	= 53.9
Spectral Subtraction Output with LPC	= 64.5

Discussion

The results of this experiment clearly show that the intelligibility of 2400 bps LPC coded speech can be significantly increased by preprocessing with spectral subtraction. These results should be considered as a lower bound for expected performance. For an actual implementation, the intermediate analog tape recording would be absent. More importantly the noise suppression algorithm could be tailored if necessary to compensate for known vocoder noise sensitivities. (This version was not tailored to operate with any specific vocoder.) Finally the noise rejection below 1kHz could be further improved by

use of an improved noise cancellation microphone.

REFERENCES

1. C. Teacher, H. Watkins, ANDVT Microphone and Audio System study, Final Report, Ketron. 1159, Aug. 1978.

2. J.S. Lim and A.V. Oppenheim, "All Pole Modeling of Degraded Speech," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-26, pp. 197-210, June 1978.

3. S.F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-27, April 1979.

4. S.F. Boll, "A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech," 1979 International Conference on Acoustics, Speech and Signal Processing, Washington, D.C., April 2-4, 1979.

5. S.F. Boll, G. Randall, R. Armantrout, R. Power, "Intelligibility and Quality Testing Results on Spectral Subtraction and LPC-10," Semi-Annual Technical Report, UTEC-CSc-80-058, May 1980.

6. W.D. Voiers, "Diagnostic Acceptability Measure for Speech Communication Systems," in Proc. Int. Conf. on Acoust., Speech, Signal Processing, Hartford, CT, May 1977, pp. 204-207.

SUPPRESSION OF ACOUSTIC NOISE IN SPEECH USING TWO MICROPHONE ADAPTIVE NOISE CANCELLATION

INTRODUCTION

It has been shown that there is a significant reduction in measured speech intelligibility and quality due to the ambient background noise generated in many operating environments [1]. A number of single microphone approaches for reducing the background noise added to speech have been developed [2]. However these methods become ineffective when the noise power is equal to or greater than the signal power or when the noise spectral characteristics change rapidly in time. This summary describes an alternative approach to noise suppression in which a second correlated noise source is adaptively filtered to minimize the output power between the two microphone signals. Three adaptive algorithm implementations were investigated: the Widrow-Hoff LMS approach [4], the lattice gradient approach [3], [4], and the frequency domain short time Fourier Transform approach [5]. Each approach was compared in terms of degree of noise power reduction, algorithm settling time, and degree of speech enhancement.

RESULTS

The performance of any noise suppression algorithm is ultimately determined by the improvement in measured intelligibility and quality due to the algorithm. Quantitative methods for measuring these improvements use scoring tests such as the DRT [6]. At the time of this experiment, a two-microphone data base was not available.

Instead a controlled data base was used to compare the performance of

these three methods: A stationary white noise source was recorded from an analog noise generator onto audio tape. The acoustic noise was generated by playing the audio tape out through a loud speaker into a hard walled room. The reference signal microphone was placed next to the loud speaker, while the primary microphone was placed twelve feet away next to the control terminal. The speaker spoke into the primary microphone while controlling the stereo recording program. The noise power was adjusted to such a level that the recorded speech was completely masked. The signals were filtered at 3.2kHz, sampled at 6.67kHz, and quantized to fifteen bits. Recordings were made with and without speech present, each lasting 24.5 sec.

For each time domain algorithm a step size was chosen such that the echo induced at the output was barely discernible. Such a choice thus represents a compromise between fast adaptation, (step size large) and minimal speech distortion, (step size small). Each algorithm then processed the acoustic data in the absence of speech activity in order to determine convergence rate versus processing time. Each method reaches a steady state error of about -15dB after about 15 seconds. Since the noise was acoustically added, no underlying clean speech spectrum was available for comparison. However, it was judged that the intelligibility of the processed speech had clearly improved. This was based upon the fact that before processing it was difficult to even detect that there was speech present in the noise, while after processing the speech was understandable.

In summary, though this two microphone approach to noise suppression requires a second signal and possibly excessive computation due to long filter lengths, it offers a potentially powerful approach for speech enhancement in

severe noise environments. Finally the processing time of frequency domain FORTRAN algorithm was approximately 3 1/2 time faster than the LMS FORTRAN algorithm as predicted due to the efficiency provided by the FFT.

REFERENCES

1. C.F. Teacher and D. Coulter, "Performance of LPC Vocoders in a Noisy Environment," in Proc. IEEE Conf. Acoust., Speech, and Signal Processing, Washington D.C., April 1979, pp. 216-219.

2. S.F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. Acoust., Speech, and Signal Processing, Vol. ASSP-27, No. 2, pp. 113-120, April 1979.

3. S. F. Boll, D. Pulsipher, "Suppression of Acoustic Noise In Speech Using Two-Microphone Adaptive Noise Cancellation," IEEE Trans. on Acoustics, Speech and Signal Proc., Dec. 1980.

4. D. Pulsipher, S.F. Boll, C.K. Rushforth, and L.K. Timothy, "Reduction of Nonstationary Acoustic Noise in Speech Using LMS Adaptive Noise Cancelling" in Proc. IEEE Conf. Acoust. Speech and Signal Processing, Wash. DC., April 1979, pp. 204-208.

5. S.F. Boll, "Adaptive Noise Cancelling in Speech Using the Short-Time Fourier Transform," in Proc. IEEE International Conf. on Acoustics, Speech, and Signal Proc., Denver, Colorado, April 9-11, 1980, Vol. 3, pp. 692-695.

TOWARDS A MATHEMATICAL THEORY OF PERCEPTION

James Kajiya

ABSTRACT

A new technique for the modelling of perceptual systems called formal modelling is developed. This technique begins with qualitative observations about the perceptual system, the so-called perceptual symmetries, to obtain through mathematical analysis certain model structures which may then be calibrated by experiment. The analysis proceeds in two different ways depending upon the choice of linear or nonlinear models. For the linear case, the analysis proceeds through the methods of unitary representation theory. It begins with a unitary group representation on the image space and produces what we have called the fundamental structure theorem. For the nonlinear case, the analysis makes essential use of infinite-dimensional manifold theory. It begins with a Lie group action on an image manifold and produces the fundamental structure formula.

These techniques will be used to study the brightness perception mechanism of the human visual system. Several visual groups are defined and their corresponding structures for visual system models are obtained. A new transform called the Mandala transform will be deduced from a certain visual group and its implications for image processing will be discussed. Several new phenomena of brightness perception will be presented. New facts about the Mach band illusion along with new adaptation phenomena will be presented. Also a new visual illusion will be presented. A visual model based on the above techniques will be presented. It will also be shown how use of statistical estimation theory can be made in the study of contrast adaptation.

Furthermore, a mathematical interpretation of unconscious inference and a simple explanation of the Tolhurst effect without mutual channel inhibition will be given. Finally, image processing algorithms suggested by the model will be used to process a real-world image for enhancement and for "form" and texture extraction.

A CONSTANT PERCENTAGE BANDWIDTH TRANSFORM FOR ACOUSTIC SIGNAL PROCESSING

James E. Youngberg

ABSTRACT

This paper describes a constant percentage bandwidth transform for acoustic signal processing. Such a transform is shown to emulate behavior found in the human auditory system, making possible both the imitation of peripheral auditory analysis, and processing which is more closely linked to perception than is possible using constant bandwidth analysis.

To enable such processing, a synthesis transformation is developed which, when cascaded with the analysis transformation, provides an analysis-synthesis identity in the absence of spectral modification. Various properties of the transform pair are derived, and a filterbank analogy is used to create a basis for intuitive understanding of the transform's operation and properties.

The effects of spectral domain modification are described and shown to be related to the properties of the analysis window function.

Principles governing discrete implementation of the transform pair are discussed, and relationships are formalized which specify the sampling of the spectral domain. These relationships are shown to depend simultaneously on the analysis window function and the selectivity (or Q) of analysis. An alternative form of the synthesis is given which facilitates a more nearly optimal logarithmic sampling of the spectral frequency axis. A minimal sampling pattern is given for the spectral domain which has an overall rate equivalent to the rate necessary to sample the constant bandwidth spectral

domain.

The nature and computation of the constant-Q spectral magnitude and phase function is discussed, and three main methods are evaluated whereby the spectral phase may be unwrapped.

Fine resolution constant-Q spectrograms are presented which show clearly the properties of constant-Q analysis applied to speech.

The use of the transform pair is discussed in the solution of the perception-related problem of time scale compression and expansion of speech. Results of this experiment are discussed.

Finally, suggestions for further research and applications are presented.

ACOUSTIC SIGNAL PROCESSING IN THE CONTEXT OF A PERCEPTUAL MODEL

Tracy Lind Petersen

ABSTRACT

The perceptual analysis of acoustic waveforms by the auditory system involves both mechanical and neural transformations of the stimulating signal. Therefore, a distinction exists between the stimulus space as characterized by acoustic vibration, and the auditory perceptual space as characterized by perceptually transformed acoustic signal information. This dissertation explores acoustic signal processing within the domain of auditory perception, beginning with the formal development of an integral transformation which can simulate certain frequency selective properties of the auditory system.

A parameterized family of analysis-synthesis transform pairs which behave as identities in the absence of perceptual modification is developed from a property of homogeneous functions. A particular member of the transform family is then implemented to simulate frequency selective properties of the peripheral auditory system. Frequency sensitivity typically found in fibers of the auditory nerve is also modeled.

Following this, an ability of the auditory brain to suppress the perception of background noise is simulated, based on a mathematical model of loudness perception. This method of noise suppression, called "perceptual subtraction"' is applied to the noise suppression processing of signals corrupted by additive noise. The signal processing results give empirical support to a theory which has been put forward to explain loudness processing by the brain.

SPEECH ARTICULATION RATE CHANGE USING RECURSIVE BANDWIDTH SCALING

H. Ravindra

ABSTRACT

Speech articulation rate change is done by analyzing the speech signal into several frequency channels, scaling the unwrapped phase signal in each channel and synthesizing a new speech signal using the modified channel signals and their scaled center frequencies. It is shown that each channel signal can be modeled as the simultaneous amplitude and phase modulation of a carrier and that only scaling the phase modulating signal does not result in a proportional scaling of the bandwidth of the channel signals which results in the introduction of different types of distortions like frequency aliasing between channels when an increase in the articulation rate is attempted and reverberation when a rate reduction is attempted. It is proposed that the amplitude modulating signal bandwidth should also be scaled and a recursive method to do this is discussed.

ESTIMATION OF THE PARAMETERS OF AN AUTOREGRESSIVE PROCESS IN THE PRESENCE OF ADDITIVE WHITE NOISE

William J. Done

ABSTRACT

Applications of linear prediction (LP) algorithms have been successful in modeling various physical processes. In the area of speech analysis this has resulted in the development of LP vocoders, devices used in digital speech communication systems. The LP algorithms used in speech and other areas are based on all-pole models for the signal being considered. With white noise excitation to the model, the all-pole LP model is equivalent to the autoregressive (AR) model.

With the the success of this model for speech well established, the application of LP algorithms in noisy environments is being considered. Existing LP algorithms perform poorly in these conditions. Additive white noise severely effects the intelligibility and quality of speech after analysis by an LP vocoder.

It is known that the addition of white noise to an AR process produces data that can be described by an autoregressive moving-average (ARMA) model. The AR coefficients of the ARMA model are identical to the AR coefficients of the original AR process. This dissertation investigates the practicality of this model for estimating the coefficients of the original AR process. The mathematical details for this model are reviewed. Those for the autocorrelation method LP algorithm are also discussed.

Experimental results obtained from several parameter estimation

techniques are presented. These methods include the autocorrelation method for LP and a Newton-Raphson algorithm which estimates the ARMA parameters from the noisy data. These estimation methods are applied to several AR processes degraded by additive white noise. Results show that using an algorithm based on the ARMA model for the data improves the estimates for the original AR coefficients.

APPLICATION OF NONPARAMETRIC RANK-ORDER STATISTICS TO ROBUST SPEECH ACTIVITY DETECTION

B. V. Cox

ABSTRACT

This report describes a theoretical and experimental investigation for detecting the presence of speech in wideband noise. A robust algorithm for making the silence-voiced-unvoiced decision is described. This algorithm is based on a nonparametric statistical signal-detection scheme that does not require a training set of data and maintains a constant false-alarm rate for a broad class of noise inputs. The nonparametric decision procedure is the multiple use of the two-sample Savage T statistic. The performance of this detector is evaluated and compared to that obtained by manually classifying twenty recorded utterances with 39, 30, 20, 10, and 0 decibel signal-to-noise ratios. In limited testing, the average probability of misclassification is less than 6 percent, 12 percent, and 46 percent for signal-to-noise ratios of 39, 20, and 0 decibels respectively.

SECTION IV

CRITICAL BAND ANALYSIS-SYNTHESIS

Tracy L. Petersen
Steven F. Boll

ABSTRACT

The formal derivation of an integral transformation which can simulate certain frequency selective (critical bandwidth) properties of the auditory system is given. A parameterized family of analysis-synthesis transform pairs which behave as identities in the absence of perceptual modification is developed from a property of homogeneous functions. The formulation facilitates a flexible choice of analysis frequencies and frequency selective response characteristics. A particular member of the transform family is then implemented to simulate frequency selective properties of the peripheral auditory system.

INTRODUCTION

Motivation

A motivation for the work presented here is based in the distinction which exists between signal representation in the stimulus domain as characterized by acoustical vibration and signal representation in the perceptual domain as characterized by the firing of neurons within portions of the auditory system. The work to be described provides a signal processing framework for modeling certain perceptually significant properties of the auditory system. It is known that the ear has bandwidth sensitivity which increases with frequency. These frequency dependent bandwidth are called critical bands and their existence has been firmly established[1]. Frequency selective characteristics of the auditory periphery are modeled through the design of an integral transform. Formulation of the transform provides a

flexible choice of analysis frequencies and frequency selective response characteristics. A particular frequency response characteristic is formulated and implemented to model the prototypical frequency sensitivity of auditory nerve fibers.

A PARAMETERIZED FAMILY OF CONSTANT-Q TRANSFORMS

Introduction

It is known that auditory critical bandwidth increases with frequency. Kajiya [2] derived a transform which is "constant-Q" in the sense that each bandpass filter involved in the transformation has a bandwidth which is a constant percentage of its center frequency. The transform Q is given by the ratio of center frequency to bandwidth. This transformation has been demonstrated as a powerful visual modeling and image processing tool [2], and also as an acoustic signal processing tool for the time stretching of speech [3].

The constant-Q transform provides a transformation integral which is similar in form to that of the short-time Fourier transform. For purposes of comparison it will be recalled that the short-time Fourier analysis integral is

$$\tilde{F}(w,t) = \int f(T)\, h(t-T)\, \exp(-jwT)\, dT \qquad (1)$$

where f(t) is the time signal to be analyzed and h(t) is the impulse response of a low-pass function. The Constant-Q analysis integral derived by Kajiya is of the form

$$F(w,t) = \int f(T)\, h[(t-T)\, w]\, \exp(-jwT)\, dT. \qquad (2)$$

The argument of the low-pass function h of equation 1 has been modified in

equation 2 to have a dependence upon frequency w. Equation 2 may be interpreted in light of a filterbank analogy. The right side of equation 4-2 may be rewritten as

$$\exp(-jwt) \int f(T)\, h[(t-T)\, w] \exp[j(t-T)\, w]\, dT$$

which means

$$F(w,t) = \exp(-jwt)\ \{f(t) * [h(wt)\ \exp(jwt)]\}. \qquad (3)$$

Thus for given w, F(w,t) is seen to be a baseband demodulation of the signal that results from convolving f(t) with a filter whose impulse response is h(wt) exp(jwt). Noting the frequency response of this filter as $\hat{H}(w,v)$ with v representing Fourier frequency, and designating F(v) as the Fourier transform of f(t), allows equation 3 to be written as

$$F(w,t) = \exp(-jwt)\ F(v)\ \hat{H}(w,v)\exp(jvt)dv/2\pi \qquad (4)$$

Homogeneous Function Formulation

A function $G(x_1,x_2,\ldots,x_n)$ is called homogeneous of degree p if for all real c>0,

$$G(cx_1,cx_2,\ldots,cx_n) = c^p[G(x_1,x_2,\ldots,x_n)].$$

Let G(w,v) be homogeneous of degree p and a bandpass function over frequency v with center frequency w. Then it can be shown[4] that the bandpass function G is Constant-Q

Analysis-Synthesis Derivation

In the constant-Q analysis integral of equation 3, the constant-Q filter function is represented in the time domain as an impulse response h(wt). In order to achieve the design flexibility which would allow the modeling of prototypical auditory filter characteristics, the frequency domain representation of equation 4 will be taken as a starting point. The relative frequency spacing of simulated auditory filters is developed in terms of position frequencies which are indicated as functions of w. For parameter p

and function R to be defined below, the bandpass frequency response $\hat{H}(w,v)$ from equation 4 is modified to be of the form

$$H_p[R_p(w),v]$$

for frequency variable v and center or position frequency $R_p(w)$, With these modifications, equation 4 becomes

$$F[R_p(w),t] = \exp[-jR_p(w)t] \quad X \qquad (5)$$

$$F(v)\ H_p[R_p(w),v]\ \exp(jvt)\ dv/2\pi$$

where F(v) is the Fourier transform of the input signal f(t), and $F[R_p(w),t]$ is the constant-Q transform of f(t) evaluated at frequency $R_p(w)$ and time t.

We now determine functions R_p, H_p, such that f(t) is recoverable from $F[R_p(w),t]$. The following lemmas and theorem are stated without proof. Their proofs can be found in [4]

LEMMA 1.

Suppose $R_p(w) = \begin{cases} \exp(w), & p=0 \\ w^{1/p}, & p>0 \end{cases}$

and $d(p) = \begin{cases} 0, & p\ R>0 \\ -\infty, & p=0. \end{cases}$

Further, suppose $H_p[R_p(w),v]$ has the following properties:

1. $H_p[R_p(w),v] = 0$ for $v \leq 0$, for all w.
2. $H_p[R_p(w),v]$ is continuous in v.
3. $H_p[R_p(w),v]$ is homogeneous of degree-p, i.e., for $c>0$

$$H_p[cR_p(w),cv] = c^{-p}\ H_p[R_p(w),v].$$

Then for every $p \geq 0$ there exists a constant B_p such that

$$I_p(v) = \int_{d(p)} H_p[R_p(w),v]dw \qquad (6)$$

$$= \begin{cases} B_p, & v>0 \\ 0, & v\leq 0. \end{cases}$$

LEMMA 2.

If p, I_p, B_p are as above, and if f(t) is a real time signal with Fourier transform F(v) such that F(0 =0, and B_p is finite, then

$F(t)=(2/B_p)$ RE[F(v) I_p(v) exp(jvt) dv/2π]

where RE(x) is the real part of x.

THEOREM.

Let f(t) be a real time signal with Fourier transform F(v) and constant-Q transform F[g,t]. Further, assume F(v)=0, v=0. Then for B_p as in Lemma 2, f(t) is recoverable from F[R_p(w),t] by the transformation

$$f(t) = (2/B_p)\ RE\{ \int F[R_p(w),t]\ \exp[jR_p(w)t]\ dw\} \qquad (7)$$

The synthesis expression of equation 7 shows that all channel signals are first modulated back to their original position frequencies after which all channel signals are integrated or summed. The real part of this sum is then scaled by the constant $2/B_p$ to recover f(t).

Critical Band Transform

In this section the constant-Q transform is first sampled in frequency and then modified to a form which approximates the critical band filterbank properties of the auditory periphery.

From lemma 1, property 3), the bandpass function H_p must be homogeneous of degree p. A function H_p which satisfies auditory filter characteristics given by Evans and Wilson [5] and also conforms to the above conditions for homogeneity and has a Q of 6 has been implemented in this work as a modified

form of the beta density function. The parameters a,b, in the following expression for H_p, were fixed experimentally to set both the Q and the skirt slopes of the filter. For Fourier frequency variable v, position frequency $R_p(w)$, and parameters a,b,

$$H_p\{R_p(w),v\} = \begin{cases} v^a\{[(b+a)/a]\ R_p(w)-v\}^b \\ (b/a)^b\ R_p(w)^{a+b+p} \\ \text{for} \quad 0<v<(b+a)/a\ R_p(w) \\ \\ 0, \text{ otherwise.} \end{cases} \tag{8}$$

In a discrete implementation, a finite set of position frequencies may be determined by evaluating $R_p(w)$ for discrete values of w. Based on the data of Wever [6], Zwislocki [7] derived a relationship between critical bandwidth and the density of neurons which connect with sensory cells of the inner ear, located along the basilar membrane. This relationship suggests that 1300 neurons approximately correspond to an interval of one critical band, and that critical bands represent uniform distance increments along the basilar membrane.

Uniform spacing on the basilar membrane corresponds to an exponential spacing of frequency measured in Hertz [8]. Thus, the position frequency function

$$R_p(w) = \exp(w)$$

is chosen which, from lemma 1, property 4), gives p=0. Discrete position frequencies of filters in the constant-Q filterbank are then given by the set

$$R_0(w_i) = \exp(w_i),\ i=1,N.$$

where

$$w_i - w_{i-1} = (w_N - w_1)/(N-1).$$

Substituting these discrete values of $R_0(w)$ into equation 5 gives

$$F[\exp(w_i),t] = \exp[-j \exp(w_i)t]$$
$$X \quad \int F(v)\, H_0[\exp(w_i),v]\, \exp(jvt)\, dv/2\pi$$
$$i=1,N, \qquad (9)$$

which specifies the constant-Q transform at the N analysis frequencies $\exp(w_i)$, i=1,N.

For this implementation, total signal bandwidth was limited to 4 kHz. Position frequencies were initially selected over 50 positions from $\exp(w_1)$=40 Hz to $\exp(w_{50})$=3900 Hz,

Because the Q of critical bandwidth drops off toward lower frequencies, the wider bandwidths in this frequency region may be achieved by summing small groups of filters from the constant-Q bank. By interactively summing groups of low frequency filters and measuring the resulting bandwidth, the 50 filters of the constant-Q bank described above were reduced to only 23 filters which closely conform to critical bandwidths. The resulting critical band filterbank is plotted in figure 1, where filters 1 through 6 have been normalized to 1. It can be shown [4] that summing these filters results in an overall frequency response which has a passband ripple of 0.2dB.

SUMMARY

Through the design of transformations which relate acoustic signals to their critical band representations, we create a means for relating signal modifications to perceptual criteria. Thus signal processing in the critical band domain may be evaluated in the stimulus domain through the combined process of reconstruction and listening to the processed signal. Additional work in the processing of critical band signals has been conducted by the

authors[9] where time-varying modifications to critical band intensities are performed to improve perceived signal-to-noise ratios.

1. D.D. Greenwood , "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Amer., Vol. 33, 1961, pp. 1344-1356.

2. J.T. Kajiya, Toward a mathematical theory of perception, PhD dissertation, University of Utah, 1979.

3. J.E. Youngberg , "Rate/pitch modification using the Constant-Q transform," Record ICASSP-79, 1979, pp. 748-751.

4. T.L. Petersen, "Acoustic Signal Processing in the Context of a Perceptual Model," Tech. report UTEC-CSc-80-113, University of Utah, June 1980.

5. E.F. Evans and J.P. Wilson, The frequency selectivity of the cochlea, Academic Press, London, 1973.

6. E.G. Wever, Theory of Hearing, Wiley, New York, 1949.

7. J. Zwislocki, Analysis of some auditory characteristics, Wiley, New York, 1965.

8. D.D. Greenwood , "Approximate calculation of the dimensions of traveling-wave envelopes in four species," J. Acoust. Soc. Amer., Vol. 34, 1962, pp. 1364-1369.

9. T.L. Petersen, S.F. Boll, "Acoustic Noise Suppression in the Context of a Perceptual Model," Proc. of the 1981 International Conference on Acoustics, Speech and Sitnal Processing, Atlanta,GA., March 1981, pp. .

ACOUSTIC NOISE SUPPRESSION IN THE CONTEXT OF A PERCEPTUAL MODEL

Tracy L. Petersen
Steven F. Boll

ABSTRACT

An acoustic noise suppression algorithm has been developed which suppresses noise from speech by first filtering it into a set of signals which approximate the loudness components perceived by the auditory system. These signals are generated by passing the input stimulus waveform through a filter bank with frequency bandwidths which approximate the ear's critical bandwidths. The noise on each signal is then suppressed using spectral subtraction techniques in a domain of simulated perception. This approach to noise suppression retains the intelligibility produced by spectral subtraction methods while eliminating the accompanying musical quality.

INTRODUCTION

The work to be described explores acoustic signal processing within the domain of perception. Such an approach requires both a knowledge of auditory system signal processing transformations, and adequate techniques for their simulation. Given a capability to map acoustic signals into the domain of perception and process this transformed information to suppress perceived levels of background noise, the processing must be followed by inverse transformations which return perceptually processed signals to an acoustic signal representation. This approach is initiated from a signal processing framework which is based on a mathematical model of peripheral auditory frequency analysis. Mathematical formulations for loudness perception and the selective identification of a tone in noise are implemented to suppress noise loudness as the simulated function of auditory brain activity. The brain's ability to concentrate upon signal components while ignoring the loudness of

background noise is described as an operation of selective listening. Each stage of the mathematical modeling is invertable. Thus it is possible to estimate processed signal intensities which in theory simulate the perception of signal loudness without imposing a need upon the brain to invoke the operation of selective listening in order to suppress the loudness of a masking background noise.

PERCEPTUAL SUBTRACTION OF NOISE

Critical Band Filtering

Peripheral auditory analysis of the ear may be likened to a bank of bandpass filters. The filters which form this auditory filter bank are called critical bands [1]. In this work we use the critical band analysis-synthesis method as given in [2, 3] This method simulates the critical band frequency analysis of the auditory periphery, while an inversion formula allows this signal to be reconstructed from its critical band filter bank analysis representation. Analysis over a 4kHz bandwidth was performed with a bank of 23 critical band filters.

Auditory Threshold and Masking

In audition the term "masking" is used to describe the situation where the loudness of a particular sound partially or completely obscures from perception a second sound. The masking sound is said to induce a threshold shift in signal detectability.

It is known that the threshold intensity of a pure unmasked tone varies as a function of tone frequency. Some workers have suggested [4] that the frequency dependent threshold shifts outside the minimum threshold region may be modeled as the result of internal masking which is inherent in the mechanisms of the auditory system itself. This approach proves to be useful

in modeling loudness perception as discussed in the remaining sections of this paper.

Loudness Perception

It is known that strong compressional mechanisms within the auditory system transform a stimulus intensity range of roughly twelve orders of magnitude down to a subjective range of approximately three or four orders. Stevens has shown [5] that loudness perception tends to be a specific mathematical function of stimulus intensity. If loudness is designated L, and stimulus intensity I, then

$$L = b\, I^{\theta} \tag{1}$$

Equation 1 gives the relationship which Stevens called the psychophysical power law. It shows loudness to be a simple power function of intensity. Hellman and Zwislocki [6] determined a value of the exponent to be 0.27.

A Model for Selective Listening

It is important to note that the critical band is an interval over which the ear integrates energy. Threshold elevation induced by an external masking noise is proportional to the noise energy within the critical band associated with the masking [7]. Zwislocki [4] has formulated an expression for loudness perception over critical band intervals which puts an additional interpretation upon the power law described in the previous section. Zwislocki reasoned that loudness perception could be represented mathematically in terms of the phenomenon of selective listening which is implicit in psychophysical masking experiments. Selective listening refers to the ability of a listener to selectively observe either the loudness of signal and noise, the loudness of signal, or the loudness of noise when signal and noise are presented simultaneously. It is the ability of the ear to perform

selective listening tasks that makes possible the measurement of loudness functions under masking[8]. Zwislocki theorized that selectively listening to a tone in noise required a subtraction of noise loudness from total loudness within the domain of perception.

In a masking situation the critical band contains the intensity I of the signal, and the intensity E of an externally presented masking noise. Here, as discussed earlier, absolute threshold is modeled as a masked threshold shift due to an internal masking intensity. M. Scharf [9] shows that the intensity M is 4 dB above the absolute threshold for a tone at critical band center frequency. According to the power law the summed intensities produce a total critical band loudness

$$L_t = b(I+E+M)^{\theta} \tag{2}$$

where b is a constant which depends on choice of units. To obtain an expression for the loudness of the signal in noise within the critical band the selective listening hypothesis is invoked to subtract off loudness due to the masking intensities. This gives the loudness of the signal L_s to be

$$L_s = b[(I+E+M)^{\theta} - (E+M)^{\theta}] \tag{3}$$

At this point it is assumed the brain has performed its selective listening operation, and in concentrating on the signal, perceives the critical band loudness L_s.

Input/Output Transformation

What is desired now as a processing goal is a stimulus domain representation of signal intensity which would induce the perception of loudness L_s while suppressing the perception of loudness due to the external masking noise. The following is a derivation of an input/output

characteristic which yields the desired intensity. L_S is first equated with the loudness $\tilde{L}_S$ that would be produced by some unmasked stimulus of intensity J. An input/output characteristic is then derived which gives J in terms of signal intensity I, external noise intensity E, internal masking intensity M, and psychophysical power exponent Θ. Because $\tilde{L}_S$ is unmasked, the expression for $\tilde{L}_S$ in terms of J has zero external noise intensity, and by definition

$$\tilde{L}_S = b[(J+M)^{\Theta} - M^{\Theta}]. \tag{4}$$

The equality of equations 3, and 4 then gives

$$b[(J+M)^{\Theta} - M^{\Theta}] = b[(I+E+M)^{\Theta} - (E+M)^{\Theta}]$$

$$(J+M)^{\Theta} = [(I+E+M)^{\Theta} - (E+M)^{\Theta}] + M^{\Theta}$$

$$J = \{[(I+E+M)^{\Theta} - (E+M)^{\Theta}] + M^{\Theta}\}^{(1/\Theta)} - M. \tag{5}$$

This new signal intensity J is one which in theory stimulates the perception of signal loudness L_S without imposing a need upon the brain to invoke the operation of selective listening in order to suppress the loudness of the external masking noise.

SIGNAL PROCESSING IMPLEMENTATION

Critical Band Signal Generation

The processing of loudness information requires the computation of intensity for each [3, 2] critical band in the analysis transform filterbank. For this implementation each critical band filter is real, zero over negative frequencies, and therefore has a complex time response.

Given a critical band filterbank composed of N filters, the kth critical band filter operates on a real input signal f(t) to produce a complex bandpass time signal. The time varying intensity, $Z_k(t)$ within the ith critical band is taken as the square of the instantaneous amplitude of the

complex signal.

The perceptual subtraction of noise as represented by equation 5 assumes that the noise is stationary and that the expected value of noise intensity within each critical band is known. Critical band noise intensity estimates were obtained by performing critical band analysis over noise only time intervals. For critical band k the expected noise intensity E_k was determined as a long-time average of the squared instantaneous envelope.

Noise Suppression

The critical band intensity $Z_k(t)$ is due to both signal and noise. Given that $J_k(t)$ is the processed intensity at the kth critical band, equation 5 then takes the form

$$J_k(t) = \{(Z_k(t)+M_k)^{\Theta} - (E_k+M_k)^{\Theta} + M_k^{\Theta}\}^{1/\Theta} - M_k \qquad (6)$$

Equation 6 defines the process of spectral subtraction in the perceptual domain as motivated by the simulation of selective listening. Critical filterbank analysis is applied to f(t), producing N complex time signals. Instantaneous intensities $Z_k(t)$ are computed and each are processed according to equation 6 to create a new critical band intensity $J_k(t)$. The appropriate inverse operations are then performed and the N channels are summed to form the output speech.

SUMMARY AND CONCLUSIONS

The success of this work both follows from and contrasts work by others using spectral subtraction [10]. The parallel between the method of perceptual subtraction and the method of spectral subtraction is that in both cases noise estimates are locally subtracted out in a transformed signal space. In the case of spectral subtraction this transformed signal space is

the short-time Fourier spectrum. In the work presented here, this transformed signal space is the perceptual space of critical band loudness, where estimated noise loudnesses are subtracted from input signal loudness.

Typically, when noise is suppressed by a time-varying attenuation of signal frequencies, successful processing requires reasonable signal-to-noise ratios. In the case of perceptual subtraction, work by Hellman and Zwislocki [8] formally suggests why this should be so. They observed that their results parallel results obtained by Miskolczy-Fodor [11] in the measurement of loudness perception by listeners with a particular hearing loss. They found that loudness curves with noise induced threshold shifts have essentially the same form as loudness curves obtained from listeners who suffer sensorineural hearing loss (recruitment) resulting in higher than normal perception thresholds. Perceptual subtraction is formulated to pass perceptible critical band signal loudnesses and suppress critical band components which contribute only noise. Based on the observation of Hellman and Zwislocki it is possible to interpret the method of perceptual subtraction processing as a method for simulating perception deafness to noise. Clearly, the opportunity to improve time-varying signal-to-noise ratios through dynamic attenuation of signal frequencies becomes limited as the noise begins to totally overtake the signal because perceptual subtraction can only preserve signal components which have perceptible intensities. In simulating perception deafness to noise we inevitably simulate deafness to signal as well when noise completely dominates the signal.

For testing purposes, speech signals were additively combined with broadband white gaussian noise at signal-to-noise ratios from 40 to 0 dB in 10

dB increments. Each sample, so constructed, was then processed for noise suppression by the method of perceptual subtraction. In evaluating both the perceptual subtraction and spectral subtraction algorithms, listening tests revealed that when the signal-to-noise ratio of the input speech becomes less that 10 dB, the quality of processed speech decreases sharply. In all cases, however, a dramatic reduction of background noise was observed. A prominent difference in processing results of this method with spectral subtraction was found to be in the overall perceived "smoothness" with which noise is suppressed. Spectral subtraction processing produces speech with a somewhat harsher quality than that produced by perceptual subtraction. Also, spectral subtraction method tends to admit small, but nevertheless sharply perceived, occurrences of noise residual artifacts. In the case of perceptual subtraction, any remaining noise artifacts were near audible threshold and judged generally less offensive.

1. D.D. Greenwood , "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Amer., Vol. 33, 1961, pp. 1344-1356.
2. T.L. Petersen, "Acoustic Signal Processing in the Context of a Perceptual Model," Tech. report UTEC-CSc-80-113, University of Utah, June 1980.
3. T.L. Petersen, S.F. Boll, "Critical Band Analysis-Synthesis," Proc. of the 1981 International Conf. on Acoust., Speech and Signal Proc., IEEE, Atlanta, Ga., March 1981, pp. .
4. J. Zwislocki, Analysis of some auditory characteristics, Wiley, New York, 1965.
5. S.S. Stevens , Psychophysics, John Wiley and Sons, New York, 1975.
6. R.P. Hellman and J. Zwislocki , "Some factors affecting the estimation of loudness," J. Acoust. Soc. Amer., Vol. 33, 1961, pp. 687-694.
7. J.E. Hawkins and S.S. Stevens , "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Amer., Vol. 22, 1950, pp. 6-13.
8. R.P. Hellman and J. Zwislocki , "Loudness function of a 1000-Hz tone in the presence of a masking noise," J. Acoust. Soc. Amer., Vol. 36, 1964, pp. 1618-1627.
9. B. Scharf, Critical bands, Academic Press, New York, 1970, pp. 157-202.
10. S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustics Speech and Signal Processing, Vol. ASSP-27, No. 2, 1979, pp. 113-120.
11. F. Miskolczy-Fodor, "Relation between Loudness and Duration of Tonal Pulses. III. Response in Cases of Abnormal Loudness Function," J. Acoust. Soc. Amer., Vol. 32, 1960, pp. 486-492.

A SPLINE APPROACH TO SPEECH ANALYSIS/SYNTHESIS

Elaine Cohen

Department ofComputer Science
University of Utah
Salt Lake City, Utah

ABSTRACT

Vocoders based on linear models of the vocal tract such as LPC result in an inverse polynomial fit of the spectrum, and along with homomorphic vocoders require pitch estimation. Analysis/synthesis using B-spline basis functions is proposed in this preliminary paper. This is a local approximation scheme which permits a concentration of parameters in regions of greater importance and employs easily computed least square coefficients. It can be used with pitch based vocoders or as a standalone nonpitch based vocoder. An experimental system, not yet optimized using special properties of human speech, has been applied to samples of male speech, female speech, simultaneous speech with two speakers, and noisy speech. Empirical results on tape will be presented.

INTRODUCTION

Two frequently used signal analysis/synthesis techniques are linear predictive coding(5) and homomorphic filtering(7). Usually both of these methods result in an approximation to the spectrum that has uniform characteristics over the whole spectrum. Using the method proposed below in conjunction with pitch extraction can lead to a pitch based vocoder with a tightness of fit that varies with the frequency. Such a fit can be done to some extent using "selective" LPC(6), but it results in piecewise inverse polynomials over the different frequency spans will probably not match at the cut frequency to provide a continuous function, much less have any derivative continuity. In fact, such a fit would be unusual. In the proposed method continuity and a somewhat variable degree of derivative continuity can be insured, but it

affects the number of parameters required.

Homomorphic filtering was developed as a general way to separate signals and can be used to determine pitch as well as vocal tract impulse response. The method in its full generality, however, involves the computation of the complex cepstrum, a difficult problem since the phase information is not in convenient form. The spline vocoder used on the power spectrum yields a model in which the variations of the i-th parameter is based on changes in energy in the signal over the band of frequencies from w_i to w_{i+k}. If the changes are not abrupt, then the parameters should change in a nonabrupt manner also. These models are based on modelling the signal as the output of a single vocal tract, so they are less robust in situations where the hypotheses are not applicable, as in the presence of multiple speakers or a noisy environment. The spline method may be used differently in this situation. Matching can be done on the real and imaginary parts of the Fourier transform of the windowed speech, and a synthetic speech waveform generated. An initial attempt using this method with many parameters has indicated that it merits further attention. Since the phase information is implicitly calculated, the problem of unwrapping the raw data is not present, and since pitch extraction is unnecessary, the quality of the speech is impervious to the pitch of the speaker (equally well for female as well as male speech), the number of speakers, or the presence of noise.

B-SPLINES

The computations involved become feasible because of the characteristics of splines in particular and B splines in particular(1,2).

Definition 1: We say $S_k(x)$ is a polynomial spline of order k over the sequence $\{x_i\}$, $x_i \leq x_{i+1}$ and $m_i = \text{card}\{x_j : x_j = x_i\}$, if

1. it is a polynomial of degree (k-1) of $x \in (x_i, x_{i+1})$,

2. $S_k(x_i) \in C^{k-1-m_j}$.

The x_i's are called the knots.

A polynomial is simply a spline with knots of multiplicity (m_i) equal to zero. Since splines are "piecewise polynomials", they can preserve the desirable characteristics of polynomial approximation while allowing more flexibility. It has been proposed to use integrals of Walsh functions as basis functions for decomposing signals(4), but they are instances of spline functions with knots at appropriate powers of (1/2).

Definition 2: The i-th B-spline of order k, $B_{i,k}(x)$, on knot set $\{x_i\}$ is a spline with the additional properties that

1. $B_{i,k}(x) = 0$ for $x \notin [x_i, x_{i+k}]$,
2. $B_{i,k}(x) > 0$, for $x \in (x_i, x_{i+k})$,
3. one of several normalizations, the two most common being $\int B_{i,k}(x) = 1$, or $\sum B_{i,k} = 1$.

The collection of B-splines form a basis for the vector space of all splines of order k on that knot sequence. Further, the above requirements mean that there are at most k nonzero functions for any value of x. The matrices that result in applying linear combinations of B-splines to solve least square or interpolation problems are then banded of width (2k-1), and hence are more easily solvable computationally. If the knots are evenly spaced, all the $B_{i,k}(x)$ are just translations of one fixed B-spline, and they have the interesting convolutional property that $B_{i,k} * B_{p,r} = B_{q,k+r}$, and hence a spline filter acting on a spline signal yields a higher order spline with known parameters. The ideal low pass filter, the Fourier window, the triangular window, and the Parzens window are all examples of B-splines occuring in signal processing, as is any other window or filter that can be represented as a piecewise polynomial. Indeed, their versatility has caused them to be used in some Computer Aided Geometric Design systems instead of Rational polynomials.

SPLINE METHOD

While the general ideas described here can be applied to a variety of situations, we shall develop an application here that fits the signal by fitting real and imaginary parts of the Fourier transform with linear combinations of B-splines and then resyntheses the the signal using the inverse transforms of the B-splines. Frequently these inverse transforms can be tabled so that the inverse transform need not be computed.

Let s(t) be a low pass filtered version of a signal, and let $s(t_p,t) = s(t-t_p)w(t)$ be the windowed signal it is desired to aproximate. For ease of presentation we call $S(w) = F[s(t_p,t)]$, where F designates the Fourier transform.

Henceforth the k denoting the order of the spline will be omitted since it is kept the same within any application. We wish to fit S(w) in some optimal

manner by a B-spline function given by

$$S(w) = \begin{cases} \sum_{i=0}^{m} c_i B_i(w), & w > 0 \\ \sum_{i=0}^{m} c_i^* B_i(-w), & \text{otherwise,} \end{cases} \tag{1}$$

where the $c_i = x_i + j\, y_i$ are complex valued parameters. We desire to minimize the error E when defined in the standard least squares sense as follows.

$$E = \int |S(w) - S(w)|^2 \tag{2}$$

The parameters are determined by minimizing E in (2) with respect to each of the parameters, which is done by setting

$$\partial E / \partial x_i = 0; \partial E / \partial y_i = 0;\ i = 0,1,\ldots m.$$

The resulting linear equations are banded since B-splines' have supports which overlap only partially, i.e.,

$$\int B_i(w) B_p(w) = 0, \quad \text{when } p \neq i-k+2, \ldots, i+k-1.$$

Inversion and solution for the parameters is computationally easier than the polynomial case using a power basis.

We next determine the equations of the time wave form corresponding to this method of fitting the spectrum. While the inverse transform of the general basis function is rather complicated, we can develop a formulation for specific instances.

Let the knots $w_i, \ldots, w_{i+k}$ be evenly spaced with spacing D_i and

$$A_i(w) = \begin{cases} c_i B_i(w), & w \geq 0, \\ c_i^* B_i(-w), & \text{otherwise.} \end{cases}$$

Then, using the convolutional propertry of uniformly spaced B-splines and the convolutional property of Fourier transforms yields

$$F^{-1}[A_i(w)] = (D_i/2) K \left(\sin(D_i w/2) / D_i w/2\right)^k \times (a_i \cos w_{i+2} t + b_i \sin w_{i+2} t) \tag{3}$$

Thus, if all the knots have spacing D the estimate of $s(t_p, t)$ is

$$\hat{s}(t_p, t) = K (D/2) \left(\sin(Dw/2) / Dw/2\right)^k \times \sum a_i \cos(i+2)Dt + b_i \sin(i+2)Dt.$$

A decaying high order trignometric polynomial with fundamental frequency D is the resulting signal. More interesting cases occur when the spacing is nonuniform. This feature allows a much closer fit in frequency ranges selectively determined to be more important to the intelligibility of the signal, and has some resemblances to selective LPC while insuring the degree of continuity desired. The simplest case uses sections of uniformly spaced knots with different spacing in each section. The inverse transforms of the transition A(w)'s will not have the simple form derived above, but the contributions of the others to the synthetic waveform are sums of decaying trignometric polynomials with different fundamental frequencies and different rates of decay. It is postulated that these parametrized waveforms carry signal information in a form faithful to the original.

APPLICATIONS

This general class of methods has not yet been widely tested or developed. However, it has been applied to a variety of selective speech signals to test for intelligibility and faithfulness in the presence of multiple speakers, female speakers, and noise at various levels, as well as on clear speech. Figures 1-4 illustrate a sequence from a voiced signal sampled at 10 kh. In each of the figures the one on the left is from the original signal, while that on the right is from the synthetic signal. Further testing to determine good knot locations and number of parameters desirable for various applications seem worthwhile, including gaining further information about the phase of the signal.

REFERENCES

(1)C. deBoor,A Practical Guide to Splines,Springer Verlag, 1978.

(2)C. deBoor,"Least Squares Cubic Spline Approximation I - Fixed Knots," CSDTR20, Purdue University, April 1968.

(3)L. L. Horowitz, "The Effects of Spline Interpolation on Power Spectral Density," IEEE Trans.Acoust.,Speech,Signal Processing, vol. ASSP-22, 22-27, February 1974.

(4)W. R. Madych, R. D. Larson, E. F. Crawford, "Piecewise Polynomial Expansions," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-25,

December 1977.

(5)J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. of IEEE, Vol. 63, April 1975.

(6)J. Makhoul, "Spectral Linear Prediction: Properties and Applications," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-23,283-296, June 1975.

(7)A.V.Oppenheim, R.W.Schafer, and T.G. Stockham, Jr., "Nonlinear Filtering of Multiplied and Convolved Signals," Proc.of IEEE, vol. 56, 1264-1291, August 1968.

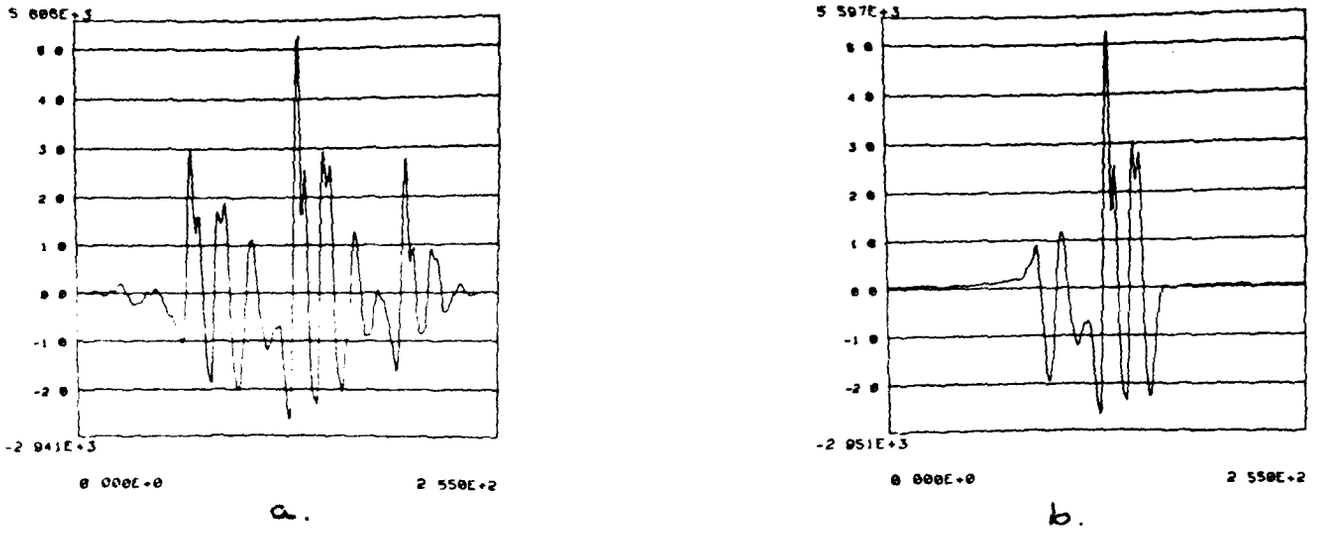

FIGURE 1: TIME SIGNAL

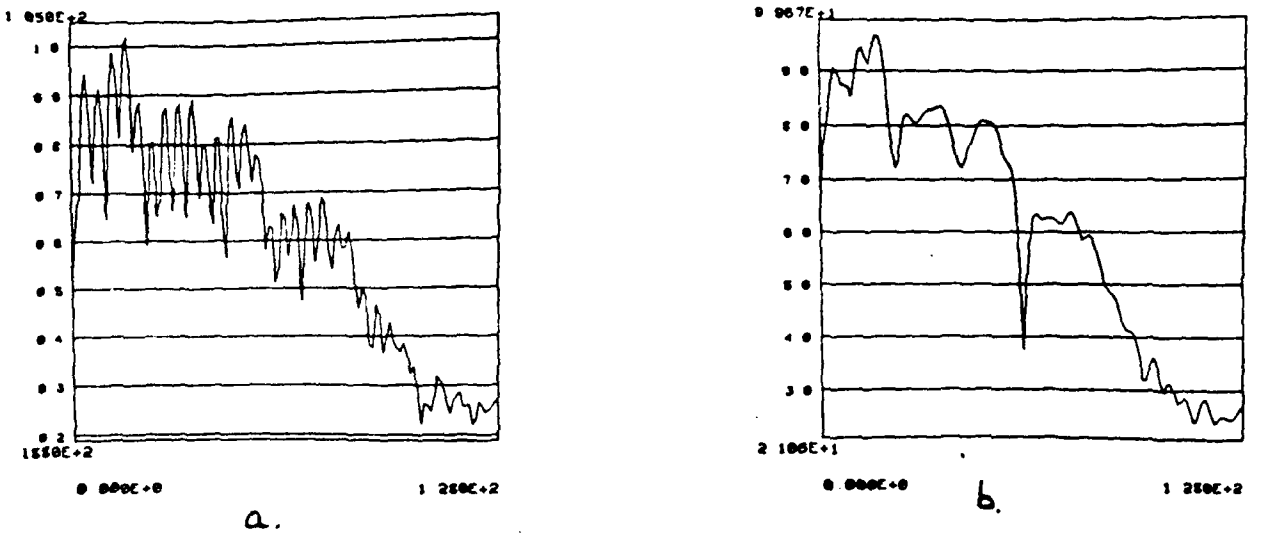

FIGURE 2: SPECTRUM (dB)

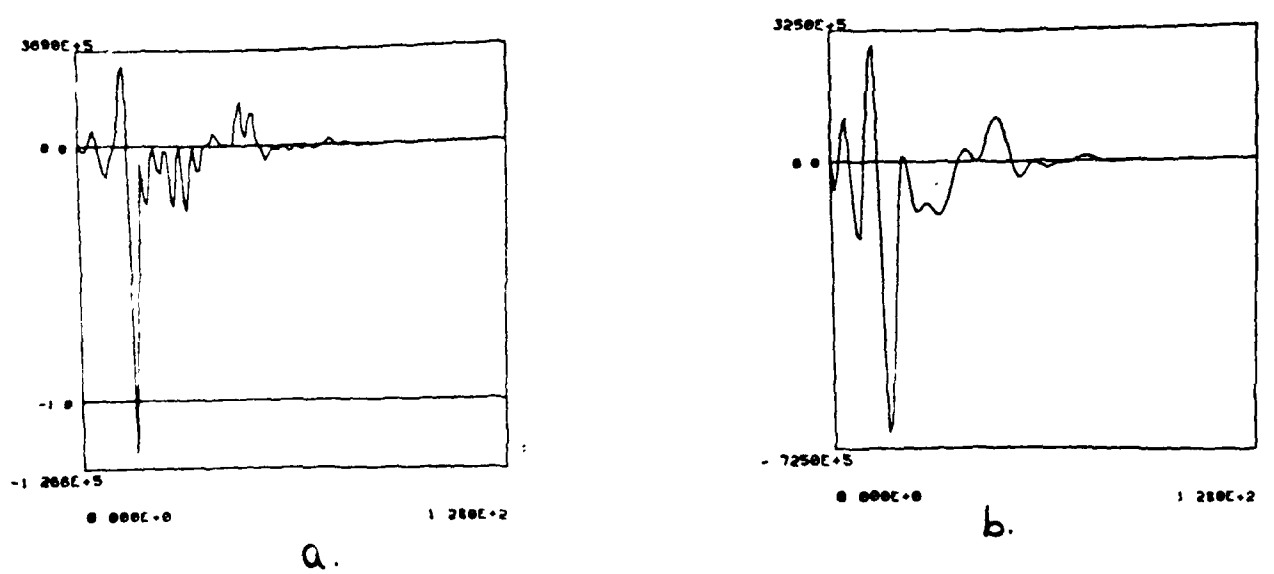

FIGURE 3: REAL PART OF FOURIER TRANSFORM

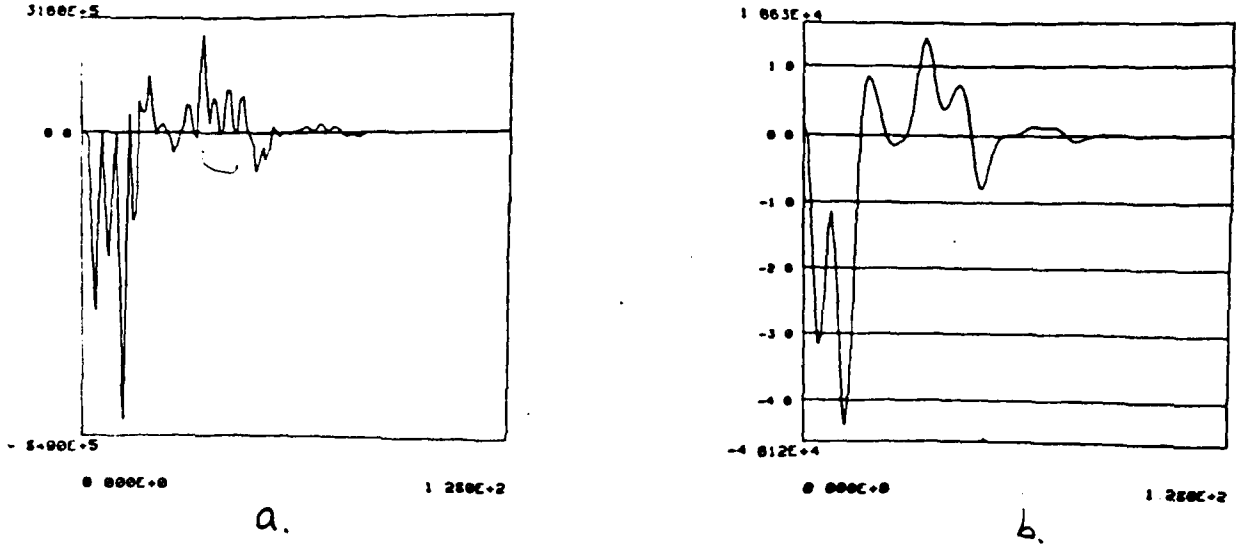

FIGURE 4: IMAGINARY PART OF FOURIER TRANSFORM

MAJOR PUBLICATIONS

[1]
S.F. Boll.
Improving Synthetic Speech Quality Using Binaural Reverberation.
ICASSP 76 1(76CH1067-8 ASSP):705-708, April, 1976.

[2]
S.F. Boll, M. Coker.
An Improved Isolated Word Recognition System Based Upon The Linear Prediction Residual.
ICASSP 76 1(76CH1067-8 ASSP):206-209, April, 1976.

[3]
S.F. Boll, W.J. Hartman.
Voice Channel Objective Evaluation Using Linear Predictive Coding.
Technical Report FAA-RD-75-189, U.S. Dept. Of Transportation, August, 1976.

[4]
S.F. Boll.
Improving Linear Prediction Analysis of Noisy Speech By Predictive Noise Cancellation.
ICASSP 77 1(77CH1197-3 ASSP):10-12, MAY, 1977.

[5]
S.F. Boll.
Noise Suppression Methods for Robust Speech Processing.
Semi-Annual UTEC-CSC-77-202, University of Utah, October, 1977.

[6]
S.F. Boll.
Noise Suppression Methods for Robust Speech Processing.
Semi-Annual UTEC-CSC-77-090, University of Utah, April, 1977.

[7]
S.F. Boll.
Noise Suppression Methods for Robust Speech Processing.
Semi-Annual UTEC-CSC-76-090, University of Utah, April, 1977.

[8]
S.F. Boll, J. Youngberg.
Constant-Q Signal Analysis and Synthesis.
ICASSP 78 1(78CH1285-6 ASSP):375-378, April, 1978.

[9]
S.F. Boll.
Suppression of Noise in Speech Using the SABER Method.
ICASSP78 1(78CH1285-6 ASSP):606-609, April, 1978.

[10]
S.F. Boll, D.C. Pulsipher, W. Done, B. Cox, L.Timothy.
Noise Suppression Methods for Robust Speech Processing.
Semi-Annual UTEC-CSC-78-204, University of Utah, December, 1978.

[11]
S.F. Boll, D. Pulsipher, W. Done, B. Cox, J. Kajiya.
Noise Suppression Method for Robust Speech Processing.
Semi-Annual UTEC-CSC-78-073, University of Utah, April, 1978.

[12]
S.F. Boll.
Suppression of Acoustic Noise in Speech Using Spectral Subtraction.
IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-27(2):113-120, April, 1979.

[13]
S.F. Boll.
A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech.
ICASSP 80 Proceedings 1(79CH1379-7):200-203, April, 1979.

[14]
S.F. Boll, D.S. Pulsipher, C.K. Rushforth, L.K. Timothy.
Reduction Of Nonstationary Acoustic Noise In Speech Using LMS Adaptive Noise Cancelling.
ICASSP 79 1(79CH1379-7):204-207, April, 1979.

[15]
S.F. Boll, D.C. Pulsipher, H. Ravindra.
Noise Suppression Methods For Robust Speech Processing.
Semi-Annual UTEC-CSC-79-163, University of Utah, October, 1979.

[16]
S.F. Boll.
Adaptive Noise Cancellation in Speech Using the Short Time Fourier Transform.
ICASSP 80 Proceedings 3(80CH1559-4):692-695, April, 1980.

[17]
S.F. Boll, R. Ravindra, G. Randall, R. Power, R. Armantrout.
Noise Suppression Methods For Robust Speech Processing.
Semi-Annual UTEC-CSC-80-058, University of Utah, MAY, 1980.

[18]
S.F. Boll, D.C. Pulsipher.
Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive noise Cancellation.
IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-28(6):752-753, December, 1980.

[19]
T. Petersen, S. Boll.
Critical Band Analysis-Synthesis.
ICASSP81 (1), March, 1981.

[20]
T. Petersen, S. Boll.
Acoustic Noise Suppression in the Context of a Perceptual Model.
ICASSP81 1, March, 1981.

[21]
T. Petersen, S. Boll.
Critical Band Analysis-Synthesis.
To be submitted to IEEE Transactions on Acoustics, Speech and Signal Processing.

[22]
T. Petersen, S. Boll.
Acoustic Noise Suppression in the Context of a Perceptual Model.
To be submitted to IEEE Transactions on Acoustics, Speech and Signal processing.

[23]
S.F. Boll, D.C. Pulsipher, W. Done, B.V. Cox, C.K. Rushforth, L. Timothy, J. Youngberg.
Noise Suppression Methods for Robust Speech Processing.
Semi-Annual UTEC-CSC-79-039, University of Utah, April, 1979.

[24]
B.V. Cox.
An Application of Nonparametric Rank-Order Statistics to Robust Speech Activity Detection.
PhD thesis, University of Utah, March, 1979.

[25]
B.V. Cox, L.K. Timothy.
Nonparametric Rank-Order Statistics Applied to Robust Voiced-Unvoiced-Silence Classification.
IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-28(5):550-561, October, 1980.

[26]
W. Done.
Estimation of the Parameters of an Autogregresive Process in the Presence of Additive White Noise.
PhD thesis, University of Utah, December, 1978.
Also Technical Report #UTEC-CSC-79--21.

[27]
W.J. Done, C.K. Rushforth.
Estimation the Parameters of a Noisy All-Pole Process Using Pole-Zero Modeling.
ICASSP-79 1(79CH1379-7 ASSP):228-231, April, 1979.

[28]
James Kajiya.
Toward A Mathematical Theory of Perception.
PhD thesis, University of Utah, March, 1979.

[29]
T. L. Petersen.
Acoustic Signal Processing in the Context of A Perceptual Model.
PhD thesis, Computer Science Dept., University of Utah, June, 1980.
Also Technical Report #UTEC-CSC-80-113.

[30]
D.C. Pulsipher.
Application of Adaptive Noise Cancellation to Noise Reduction.
PhD thesis, University of Utah, March, 1979.
Also Technical Report# UTEC-CSC-79-022.

[31]
J.E.Youngberg.
A Constant Percentage Bandwidth Transform for acoustical Signal Processing.
PhD thesis, University of Utah, June, 1979.
Also Technical Report UTEC-CSC-80-004.